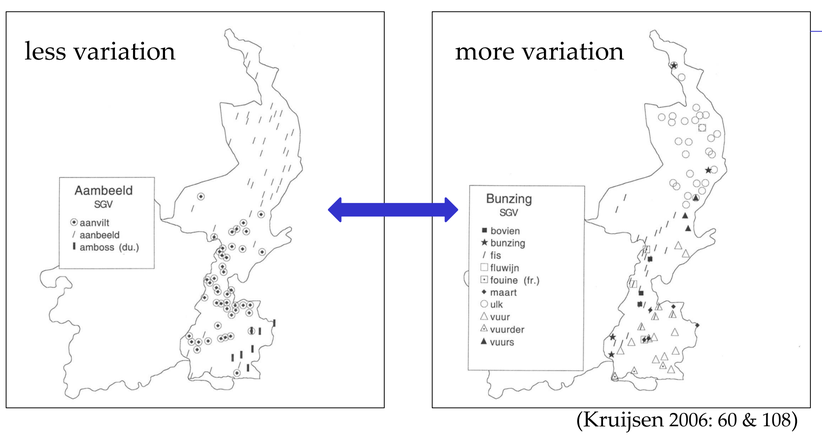
Do semantic features influence lexical geographical variation in dialect data?

Karlien Franco, Dirk Geeraerts & Dirk Speelman
Quantitative Lexicology and Variational Linguistics (QLVL)
KU Leuven, University of Leuven

Introduction



larger number of lexical types for concept BUNZING
&
variants scattered across space in less homogeneous way

(Kruijsen 2006: 60 & 108)

→ differences in amount of variation assumed to be influenced by **lectal or geographical factors** (e.g. Nerbonne & Kleiweg 2003, Séguy 1971, Wieling, Nerbonne & Baayen 2011)

pilot study (Speelman & Geeraerts 2008, Geeraerts & Speelman 2010)
data: *the human body* in the Dictionary of Limburgish Dialects
lexical dialect variation influenced by **concept characteristics** as well:
→ more lexical geographical variation for:

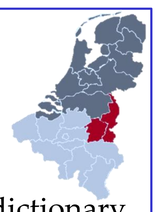
< prototype theory

- **vaguer** concepts — e.g. ACHTERWERK-ACHTERSTE 'bottom' vs. DUIM 'thumb'
- **less salient** concepts — e.g. SLUIK HAAR 'straight hair' vs. HOOFD 'head'
- concepts prone to **negative affect** — e.g. KWIJL 'drool' vs. JUKBEEN 'cheekbone'

research questions

How stable is the effect of vagueness, salience and affect in other semantic fields?
Do differences in the amount of variation occur between semantic fields?

Data: Dictionary of Limburgish Dialects



digitized database:

concept	lexical item	location	…
BRUIDEGOM 'groom'	*jongen*	Lommel	…
BRUIDEGOM 'groom'	*bruidegom*	Oirlo	…
BRUIDEGOM 'groom'	*man*	Oirlo	…
…	…	…	…
STERVEN 'to die'	*doodgaan*	Venray	…
STERVEN 'to die'	*de hoek omgaan*	Bree	…
STERVEN 'to die'	*doodgaan*	Neer	…
…	…	…	…

- onomasiological dictionary
- focus on concepts from 4 chapters:
  - *the human body*
  - *the physical and abstract world*
  - *personality and feelings*
  - *family and sexuality*

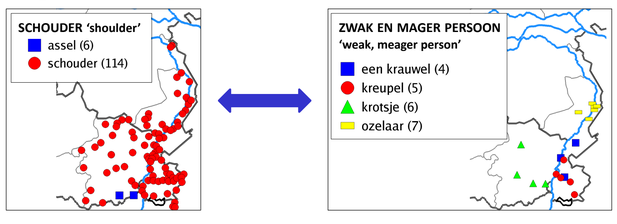
→ each chapter represents one semantic field

Methodology

linear regression using quantitative operationalization of predictors and response

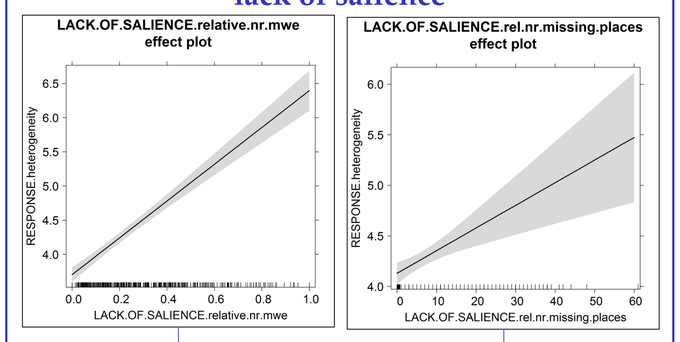
- predictors:
  - semantic field of concept (SEMFIELD)
  - LACK OF SALIENCE per concept
  - VAGUENESS per concept
  - proneness to AFFECT
- response: log(diversity * fragmentation)
  - diversity: number of lexical types per concept
    e.g. BLOED 'blood': 1 lexical item (*bloed*)
    vs. LUIEREN 'to be lazy': 27 lexical items (*faulenzen*, *lamzakken* …)
  - fragmentation: numeric estimate of amount of geographical scatter in the data per concept

→ we expect positive estimates in the regression model for all predictors (except for semantic field)



Predictors & results

lack of salience

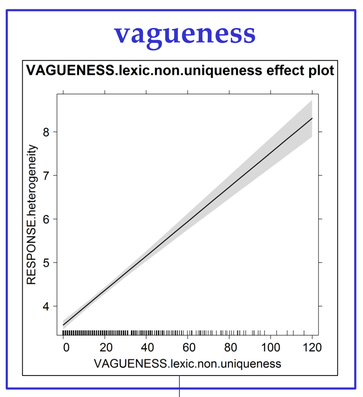


calculation: proportion of multi-word expressions per concept

calculation: proportion of places without a response per concept

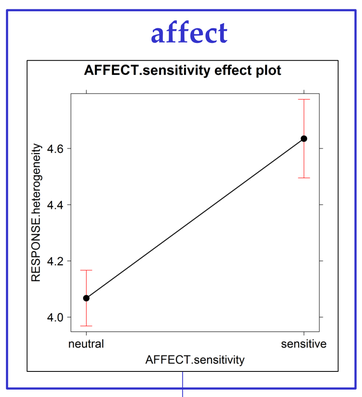
- predictors behave as expected:
  - more variation for less salient concepts
  - more variation for vaguer concepts
  - more variation for concepts that are sensitive to affect
- we also find differences between semantic fields
- Adjusted $R^2 = 0.5128$

vagueness



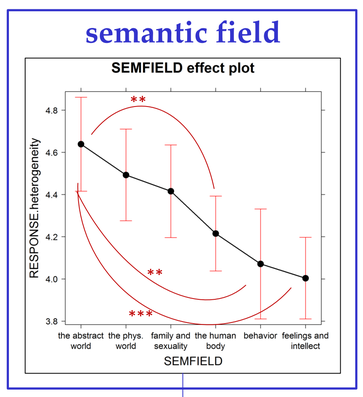
calculation: number of lexical types per concept that are used to refer to other concepts as well

affect



coding procedure: all concepts coded manually for sensitivity to affect

semantic field



coding procedure: based on onomasiological classification in dictionary

Discussion & conclusions

- The analysis provides further evidence for the importance of protoype-theoretical concept characteristics (viz. vagueness and salience) for lexical geographical variation and confirms the influence of affect sensitivity on lexical geographical variation.
- The influence of these concept features seems to be relatively stable across semantic fields, although differences between semantic fields occur:
  - the amount of lexical geographical variation is higher in *the abstract world* in comparison to *behavior* and *feelings and intellect: the abstract world* contains a smaller amount of concepts that are vague and less salient
  - the influence of concept concreteness may explain the significant difference between *the abstract world* and *the human body*, but further research is necessary

References

Geeraerts, D. and Speelman, D. 2010. Heterodox concept features and onomasiological heterogeneity in dialects. In Geeraerts, D., Kristiansen, G. and Peirsman, Y. (eds.), *Advances in Cognitive Sociolinguistics*, 23-40. Berlin / New York: De Gruyter Mouton.
Kruijsen, J. 2006. *Een eeuw lang Limburgs. SGV-enquête 1914 - Veldeke 2006*. Maastricht: TIC.
Nerbonne, J., & Kleiweg, P. 2003. Lexical Distance in LAMSAS. *Computers and the Humanities, 37(3),* 339–357.
Séguy, J. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane, 35*, 335–357.
Speelman, D. and Geeraerts, D. 2008. The role of concept characteristics in lexical dialectometry. *International Journal of Humanities and Arts Computing 2(1–2).* 221–242.
Wieling, M., Nerbonne, J., & Baayen, R. H. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PloS One, 6(9),* e23613.

Contact

karlien.franco@kuleuven.be
dirk.geeraerts@arts.kuleuven.be
dirk.speelman@arts.kuleuven.be
http://wwwling.arts.kuleuven.be/qlvl/

QITL-6
6th Conference on Quantitative Investigations in Theoretical Linguistics
5 November 2015, Tübingen